



SvPablo and the PERC Performance Tool Suite

Evan Welbourne, NERSC/LBNL
NUG Meeting, June 4th, 2002

Daniel A. Reed, Ruth A. Aydt, Luiz DeRose, Celso L. Mendes, Randy L. Ribler, Eric Shaffer, Huseyin Simitci, Jeffrey S. Vetter, Daniel R. Wells, Shannon Whitmore, and Ying Zhang, "*Performance Analysis of Parallel Systems: Approaches and Open Problems*," Joint Symposium on Parallel Processing (JSPP), Nagoya, Japan, June 1998 (invited paper and keynote presentation), pp, 239-256.

PERC consortium, *The Original Proposal for the PERC Project*.
Proposal in response to the DOE SciDAC Solicitation 01-07,
Mar. 9, 2001

PERC consortium, *The Performance Evaluation Research Center (PERC)*.
slides from the poster session at the SciDAC PI meeting,
Jan. 15-16, 2002

PERC consortium, personal contact

Slides 13-38 adapted from:

Celso Mendes, Ying Zhang, Dan Reed,
SvPablo: A Toolkit for Performance Analysis of Parallel Systems.
slides from tutorial given at the PERC SciDAC meeting during SC2001,
Nov. 2002



The Changing face of High Performance Computing



As the science and technology of high performance computing advances, our applications and the environment in which they execute is constantly evolving.

- Larger systems: HPC systems are scaling to a larger and larger numbers of processors.
 - Complex codes: Codes are often multi-language, and written using object-oriented, and data-parallel programming languages.
 - Diverse environment: With the advent of the computational grid, our codes will no longer execute on homogeneous systems, but on a large heterogeneous system.
-
- ✍ Application performance is a complex function of many variables; it's often counter-intuitive, and probably not predictable using first-principles. Our performance tools are not suited for the future of HPC.



Requirements for the Success of a Modern Performance Toolkit



To succeed in the rapidly evolving world of HPC, a performance toolkit should be:

- ✍ Scalable: Tools should scale for use on large systems.
- ✍ Portable: Tools should work on every system in the grid.
- ✍ Compatible: Tools should be work with most programming languages.
- ✍ Versatile: Tools should leverage a variety of HW and SW techniques.
- ✍ Experimental: Since analysis from first-principles is unlikely.

As always, we would like our performance tools to be:

- ✍ Easy to use: Simple and intuitive for users.
 - ✍ Non-intrusive: Disruption of the normal execution and usage pattern for an application should be minimal.
-
- ✍ The PERC project aims to develop a toolkit with these qualities.

The Performance Evaluation Research Center (PERC) is an “Integrated Software Infrastructure Center” (ISIC) sponsored under DoE’s SciDAC program.



- Funding: approx. \$2.4 million per year.
- Mission:
 - Develop a science of performance.
 - Engineer tools for performance analysis and optimization
- Focus:
 - Large, grand-challenge calculations, especially SciDAC application projects.

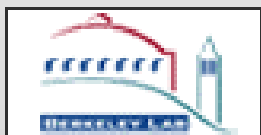
PERC website: <http://perc.nersc.gov>



PERC Participants



PERC is a collaboration among eight institutions: four DoE laboratories, and four Universities.



Lawrence Berkeley National Lab

David Bailey
Erich Strohmaier



Argonne National Lab

Paul Hovland
Boyana Norris



Lawrence Livermore National Lab

Dan Quinlan
Bronis de Supinski
Jeffery Vetter



Oak Ridge National Lab

Pat Worley
Tom Dunigan

University of Tennessee

Jack Dongarra



University of Illinois at Urbana-Champaign

Dan Reed



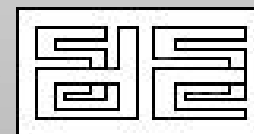
University of Maryland at College Park

Jeff Hollingsworth



San Diego Supercomputing Center

Allan Snaveley



The PERC project is led by David H. Bailey at NERSC/LBNL.

PERC's tool development effort is led by Dan Reed at UIUC.



The PERC Performance Tool Suite



The goal of PERC's tool effort is to produce an interoperable suite of measurement, analysis, and tuning tools that are suited for use on current and future HPC systems.

This goal requires three tightly coupled research efforts:

- End-user tools that integrate various analysis and measurement approaches, providing a common interface for comparing performance measurements across platforms and executions and correlating this data with benchmark and application source code.
- Flexible instrumentation systems for capturing hardware and software interactions, instruction execution frequencies, memory reference behavior, and execution overheads.
- Data management infrastructure for tracking performance experiments and data across time and space.



An API for portable hardware measurement

Provides the tool designer with a consistent interface and methodology for use of the performance counter hardware found in most major microprocessors.

PAPI is available for Linux/x86, Windows 2000, Linux/IA-64, Sun Solaris/Ultra 2.8, IBM AIX/Power, SGI IRIX/MIPS, Compaq Tru64/Alpha Ev6 & Ev67, and Cray T3E/Unicos.

To use PAPI on Seaborg, issue the command:

```
% module load papi
```

For more information on PAPI at NERSC, see the NERSC help page:

<http://hpcf.nersc.gov/software/tools/papi.html>

PAPI project homepage: <http://icl.cs.utk.edu/projects/papi/>

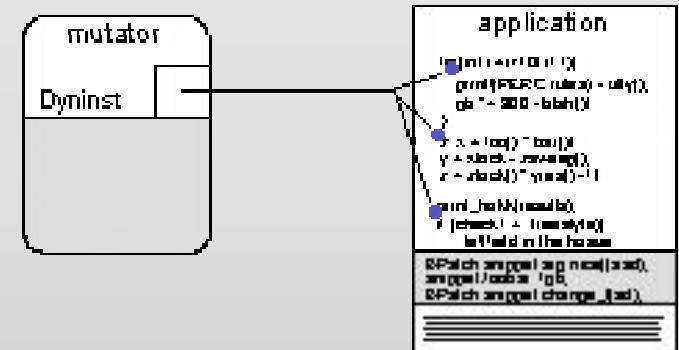


An API for dynamic instrumentation at runtime.

Provides a machine independent interface to permit the creation of tools that use runtime code patching.

The interface is analogous to a machine independent intermediate representation of instrumentation as an abstract syntax tree.

The Dyninst API is available for MIPS (IRIX), Power/PowerPC (AIX), SPARC (Solaris), and x86 (Linux, Solaris and NT).



Dyninst project homepage: <http://www.dyninst.org>



A set of tools for establishing bounds on the performance of an application or program construct.

Will use source code analysis to determine what sections of code are memory bandwidth limited, instruction scheduling limited, etc. on a given architecture.

The tools will also utilize optional user annotations in order to provide more accurate bounds for performance-critical sections of code

A tool for memory hierarchy measurement.

Uses runtime instrumentation to extract a detailed representation of the memory reference pattern of an application.

The memory reference pattern information will be the input to a collection of post-execution tools that provide insight into memory performance issues such as cache conflicts and memory bandwidth contention.

The Sigma effort is a joint collaboration between IBM and the University of Maryland.



Pablo

An end-user tool that supports source instrumentation and browsing of runtime performance data with a graphical user interface.

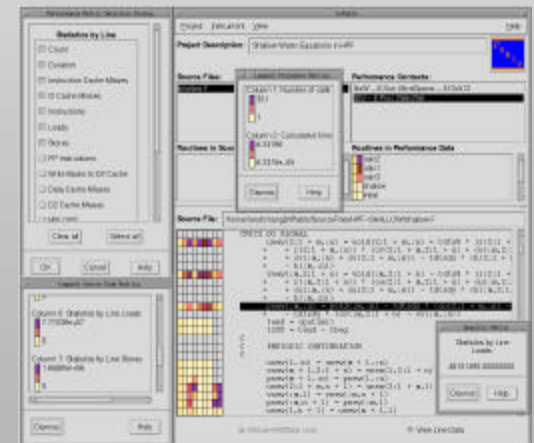
Incorporates the APIs and other performance tools to provide a front-end to the PERC tool suite.

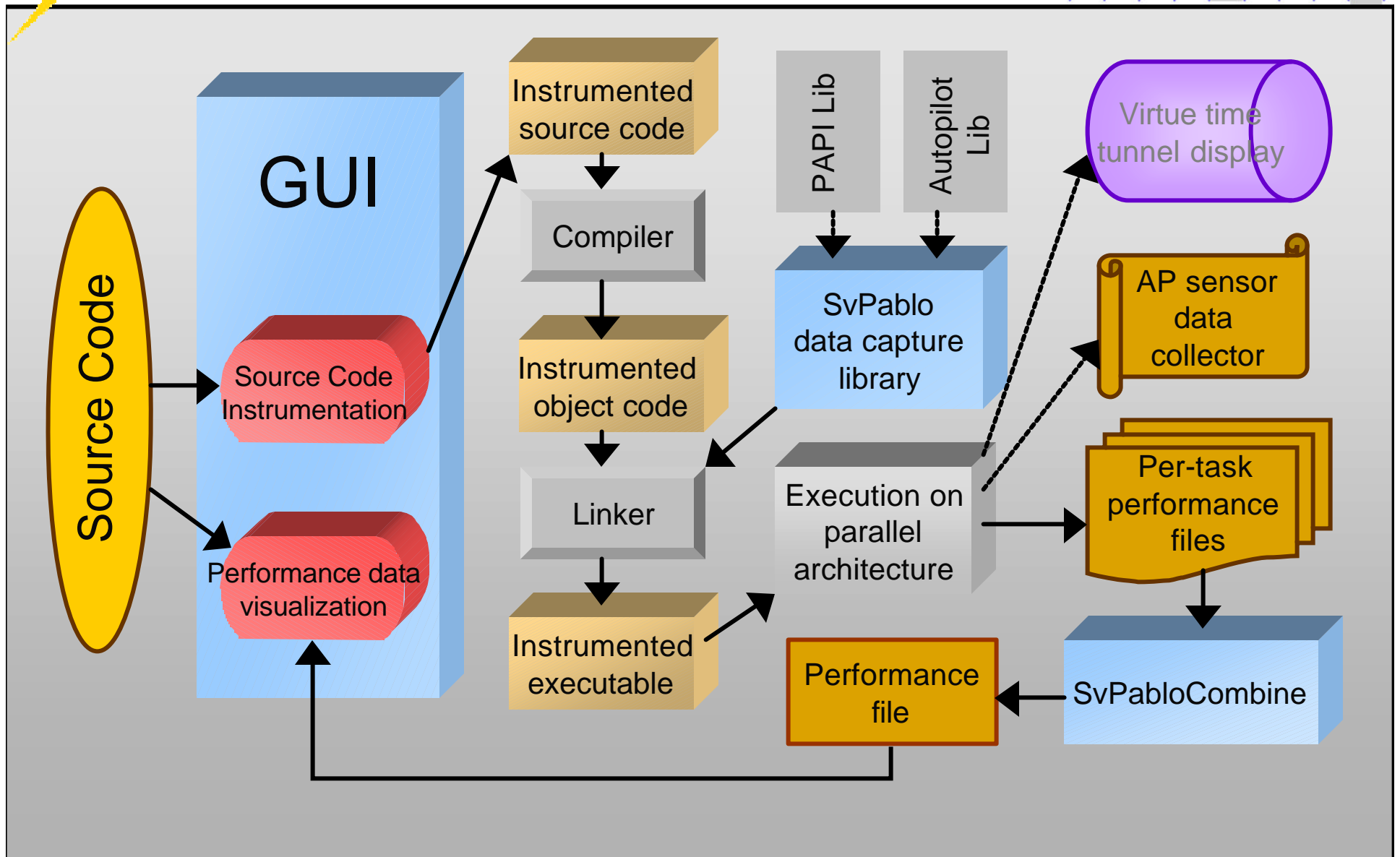
SvPablo is available for: SunOS 5.7, SGI Irix 6.5, IBM SP2 AIX 4.3, RedHat Linux 6.1, 6.2, 7.1, Intel Itanium IA-64/RedHat 7.1, and Compaq Alpha OSF1 5.1

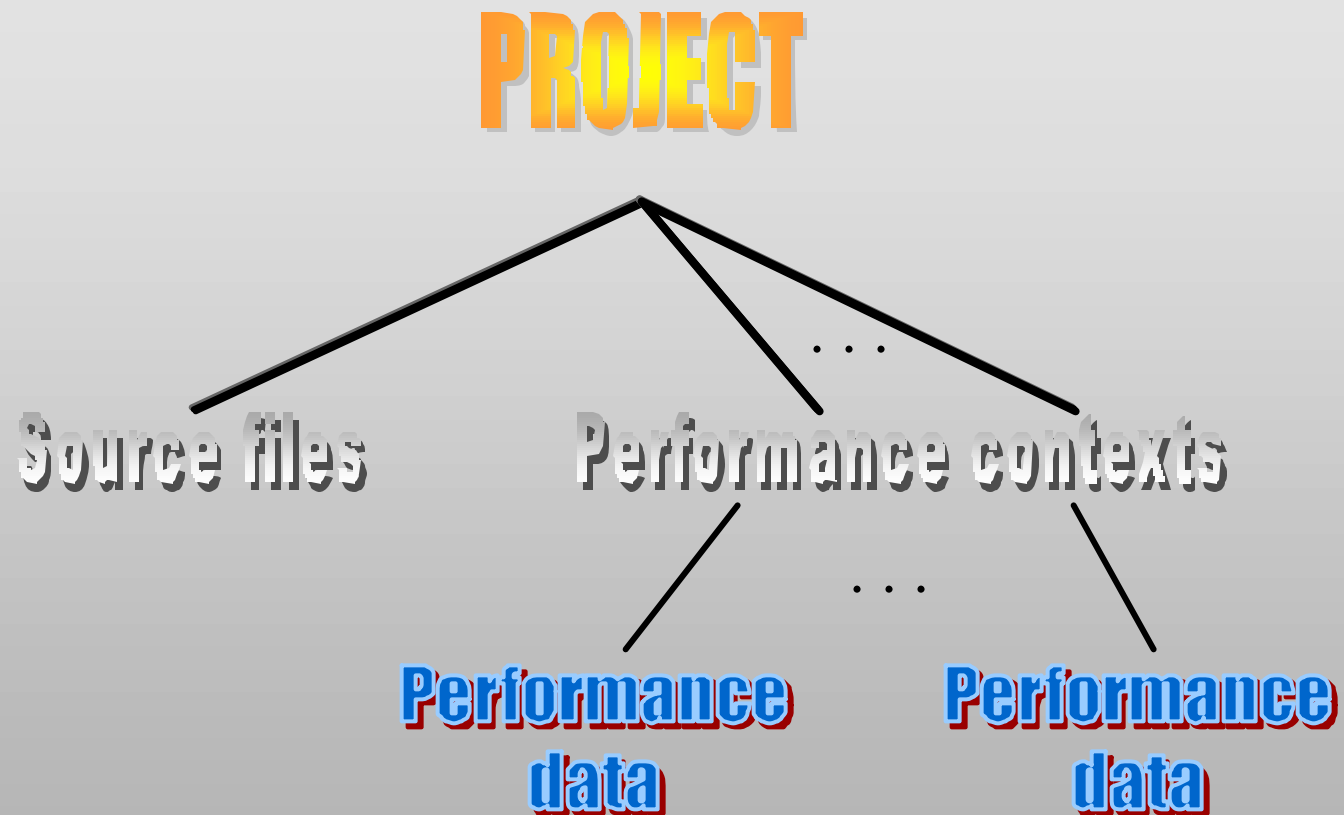
Supports C, Fortran-77, Fortran-90 (free and fixed form), HPF, and MPI

SvPablo project homepage:

<http://www-pablo.cs.uiuc.edu/Software/SvPablo/svPablo.htm>







SvPablo v5.1 is installed on NERSC's Seaborg machine.

To use SvPablo on Seaborg, issue the command:

```
% module load svpablo
```

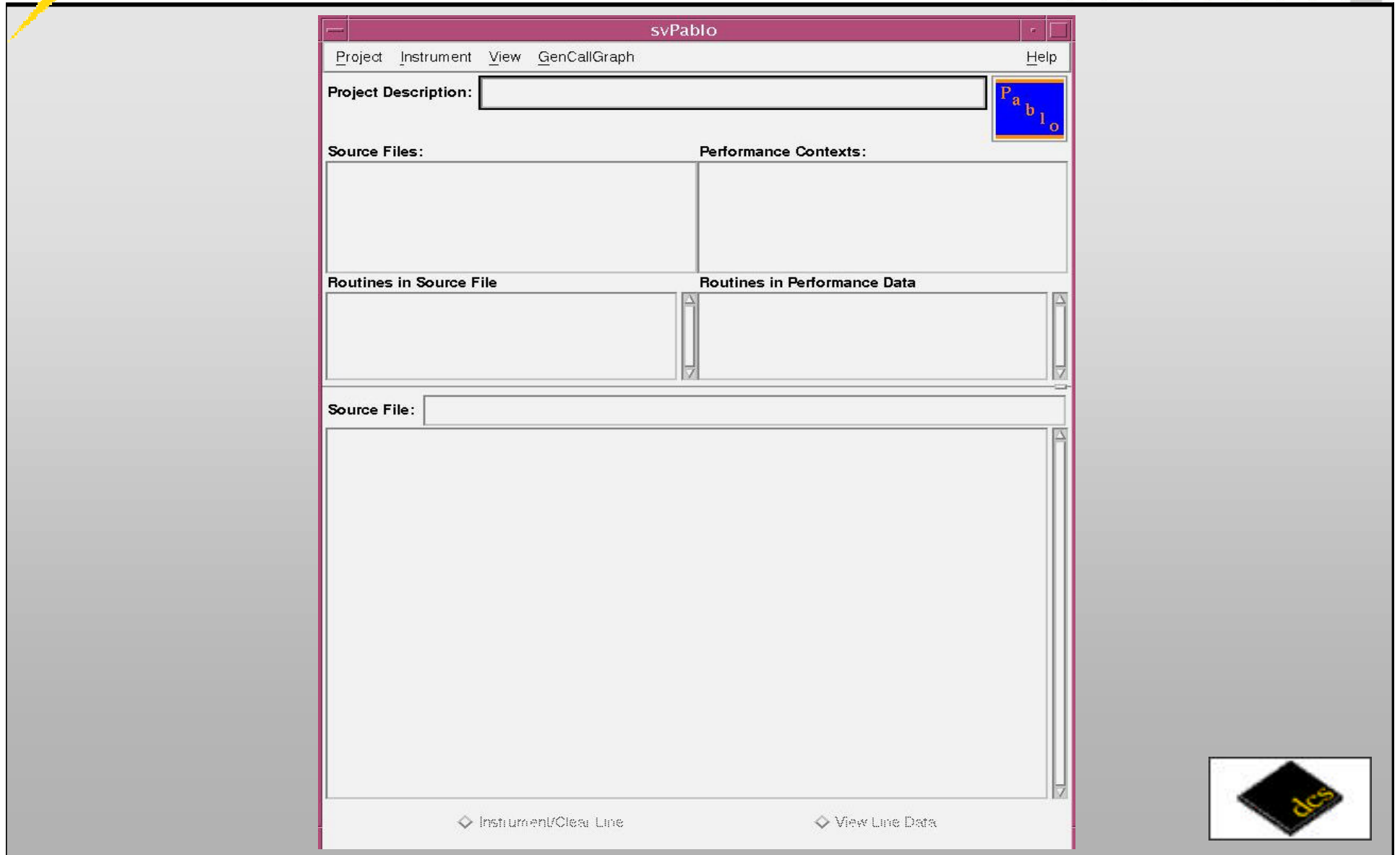
SvPablo's main window can then be launched with the command:

```
% runSvPablo &
```

The `SvPabloCombine` command may be issued on the command line as well:

```
% SvPabloCombine <parameters>
```





svPablo

Project Instrument View GenCallGraph Help

Project Description:

Source Files:

Performance Contexts:

Routines in Source File

Routines in Performance Data

Source File:

Instrument/Clear Line

View Line Data

Pablo



New Project Dialog Box

New Project

Project Directory:

o/Install/svPabloProjects/svTempProject/

Change

Project Description:

Source Directory:

d/derose/DemoPablo/Install/SourceFiles

Change

Source Files:

Add

Change

Delete

Performance Contexts:

Add

Change

Delete

Parser Options:

[C]

[Fortran 77]

[Fortran 90]

Change Parser Options...

OK

Cancel

Help



Add Performance Context

Context Description:

Context Directory:

PerformanceFile:

Instrument Directory:



Interactive Instrumentation and Analysis



Typically done in three steps:

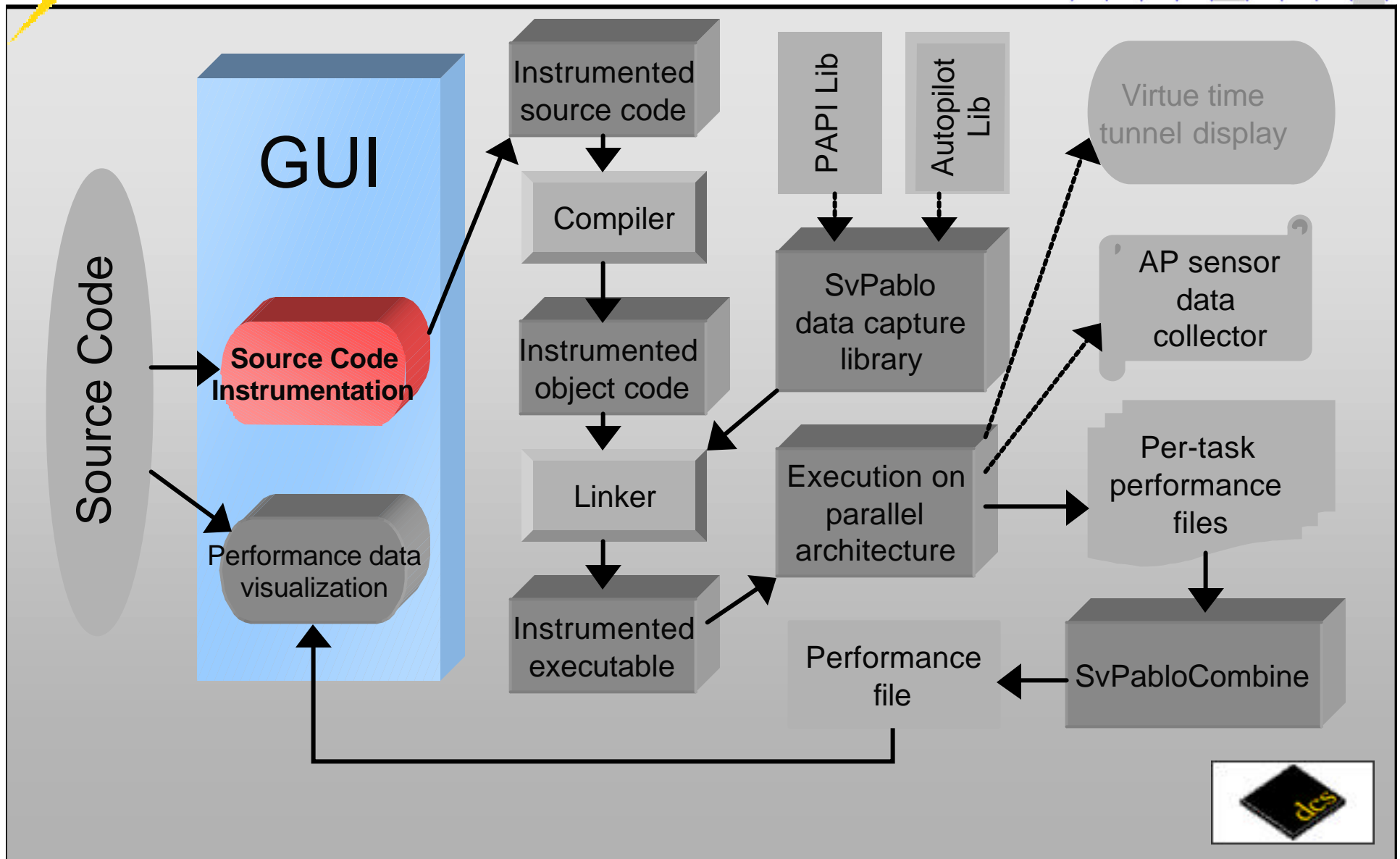
1. source code instrumentation
2. program compilation and execution
3. performance data visualization

Each cycle (1-2-3) corresponds to a Performance-Context

If desired, the cycle can be repeated (multiple performance-contexts)

Steps 1 and 3 are done in the GUI





- Instrumentable constructs:
 - ✍ procedure calls
 - ✍ outer loops
- Basic metrics:
 - ✍ counts
 - ✍ inclusive durations
 - ✍ exclusive durations
- Optional metrics:
 - ✍ any metric provided by PAPI

instrumentable
constructs
(function calls
and outer loops)

svPablo

Project Instrument View Help

Project Description: Red Black SOR in C using MPI

Source Files:

prbsor.c
prelax.c
p_io.c

Performance Contexts:

Origin 2000 - 16 R10K processors - 800x800
Power Challenge - 8 R10K Processors - 125x125
NoW - 8 Sun UltraSparcs - 800x800
NoW - 4 Sun UltraSparcs - 125x125
Example: no instrumentation

Routines in Source File

main
MPI_Comm_size
MPI_Comm_rank
MPI_Get_processor_name
fprintf

Routines in Performance Data

Source File: /home/reed/derose/DemoPablo/Install/SourceFiles/CMPI_RBSOR/prbsor.c

```

MPI_Init( &argc, &argv );
> MPI_Comm_size( MPI_COMM_WORLD, &numprocs );
> MPI_Comm_rank( MPI_COMM_WORLD, &myid );
> MPI_Get_processor_name( processor_name, &namelen );

> fprintf(stderr, "Process %d of %d on %s\n", myid, numprocs, processor_name );

if (myid == 0)
> errorExit = readInput( &n, &it, &rhs );

> MPI_Bcast(&errorExit, 1, MPI_INT, 0, MPI_COMM_WORLD);


if (errorExit > 0)
{
> printf("Processor: %d exiting due to error: %d\n", myid, errorExit);
> exit(1);
}

> MPI_Bcast(&n, 1, MPI_INT, 0, MPI_COMM_WORLD);
> MPI_Bcast(&it, 1, MPI_INT, 0, MPI_COMM_WORLD);
> MPI_Bcast(&rhs, 1, MPI_DOUBLE, 0, MPI_COMM_WORLD);

```

Instrument/Clear Line

View Line Data



Instrumentable lines

Instrumented lines

svPablo

Project Instrument View GenCallGraph Help

Project Description: MILC on IA-64

Source Files:
control.c
setup.c
update.c
d_congrad5.c
com_mpi.c

Performance Contexts:
IA-64 with 4 Processors
IA-64 with 16 Processors

Routines in Source File
main
initialize_machine
g_sync
setup
setup_analyze

Routines in Performance Data

Source File: /u/nosa/svpablo/Celso/MILC/sources/control.c

```

> initialize_machine(argc,argv);
> g_sync();
  /* set up */
> prompt = setup();
> setup_analyze();


  /* loop over input sets */
> while( readin(prompt) == 0){
  /* perform warmup trajectories */
> dttime = -dclock();
  for(traj_done=0; traj_done < warms; traj_done++ ){
~> update();
  }
> if(this_node==0)printf("WARMUPS COMPLETED\n");

  /* perform measuring trajectories, reunitarizing and measuring */
  meascount=0; /* number of measurements */

~> plp = cmplx(99.9,99.9);
  avm_iters = avs_iters = 0;

```

Instrument/Clear Line View Line Data



svPablo

Project Instrument View Help

Project Description: Red Black SOR in C using MPI

Source Files:
prbsor.c
prelax.c
p_io.c

Performance Contexts:
Origin 2000 – 16 R10K processors – 800x800
Power Challenge – 8 R10K Processors – 125x125
NoW – 8 Sun UltraSparcs – 800x800
NoW – 4 Sun UltraSparcs – 125x125
Example: no instrumentation

Routines in Source File
main
MPI_Comm_size
MPI_Comm_rank
MPI_Get_processor_name
fprintf

Routines in Performance Data

Source File: /home/reed/derose/DemoPablo/Install/SourceFiles/CMPI_RBSOR/prbsor.c

```

/*
 * The following line has statements grouped together
 * to test some functionalities of the SvPablo GUI.
 */
> for (i=i; i<=it; i++) { relax(n,&omega,f,u,&mynorm); updateOmega(&omeg
> MPI_Sendrecv( &u[myend*n], n, MPI_DOUBLE, botton, (myid+1)*blocksize
u, n, MPI_DOUBLE, top, myid * blocksize,
MPI_COMM_WORLD, &status );

> MPI_Sendrecv( &u[n], n, MPI_DOUBLE, top, myid * blocksize + 1,
&u[(myend+1)*n], n, MPI_DOUBLE, botton,
(myid+1) * blocksize + 1, MPI_COMM_WORLD, &status );

if (myid == 0) oldNorm = norm;
> MPI_Reduce(&mynorm, &norm, 1, MPI_DOUBLE, MPI_SUM, 0, MPI_COMM_WORL
}

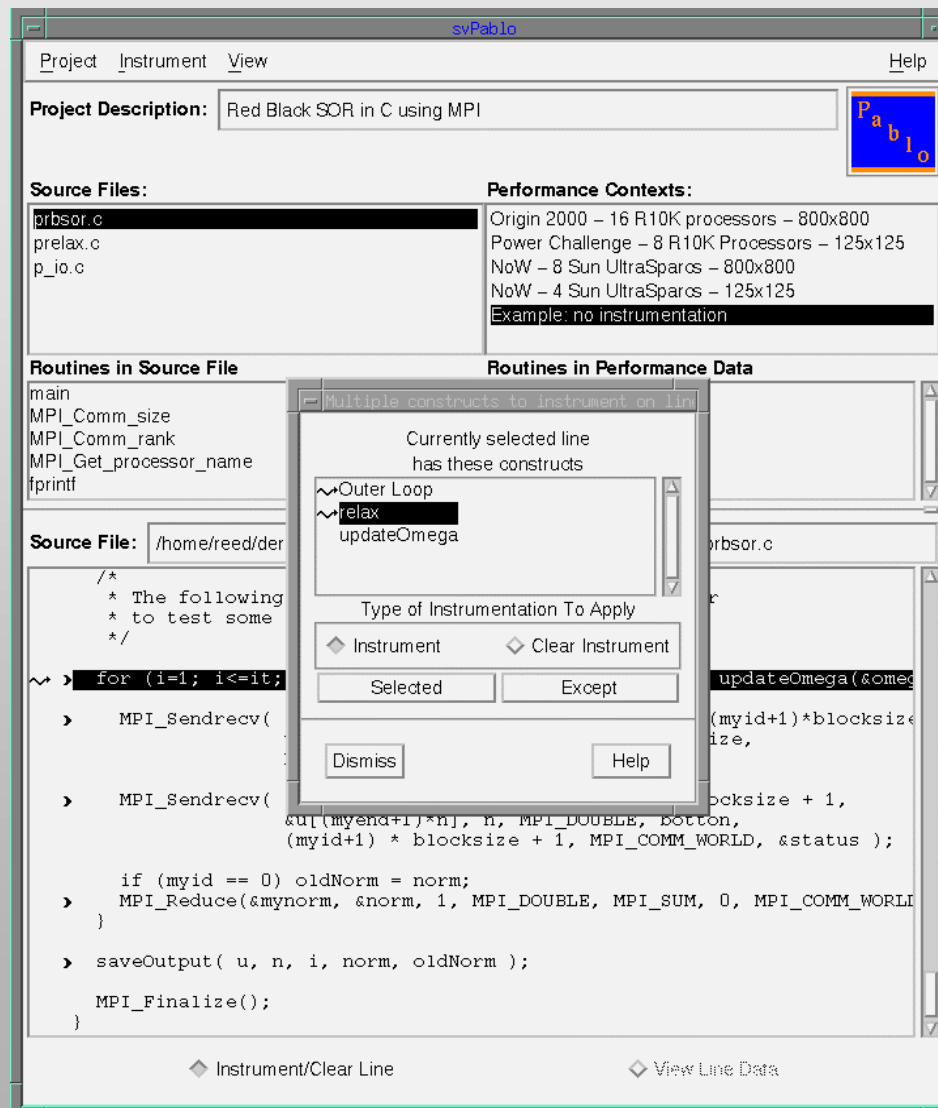
> saveOutput( u, n, i, norm, oldNorm );

MPI_Finalize();
}

```

Instrument/Clear Line View Line Data





The screenshot shows the svPablo GUI with the following sections:

- Project Description:** Red Black SOR in C using MPI
- Source Files:** prbsor.c, prela.c, p_io.c
- Performance Contexts:**
 - Origin 2000 – 16 R10K processors – 800x800
 - Power Challenge – 8 R10K Processors – 125x125
 - NoW – 8 Sun UltraSparcs – 800x800
 - NoW – 4 Sun UltraSparcs – 125x125
 - Example: no instrumentation
- Routines in Source File:** main, MPI_Comm_size, MPI_Comm_rank, MPI_Get_processor_name, fprintf
- Routines in Performance Data:** (Empty)
- Source File:** /home/reed/der...

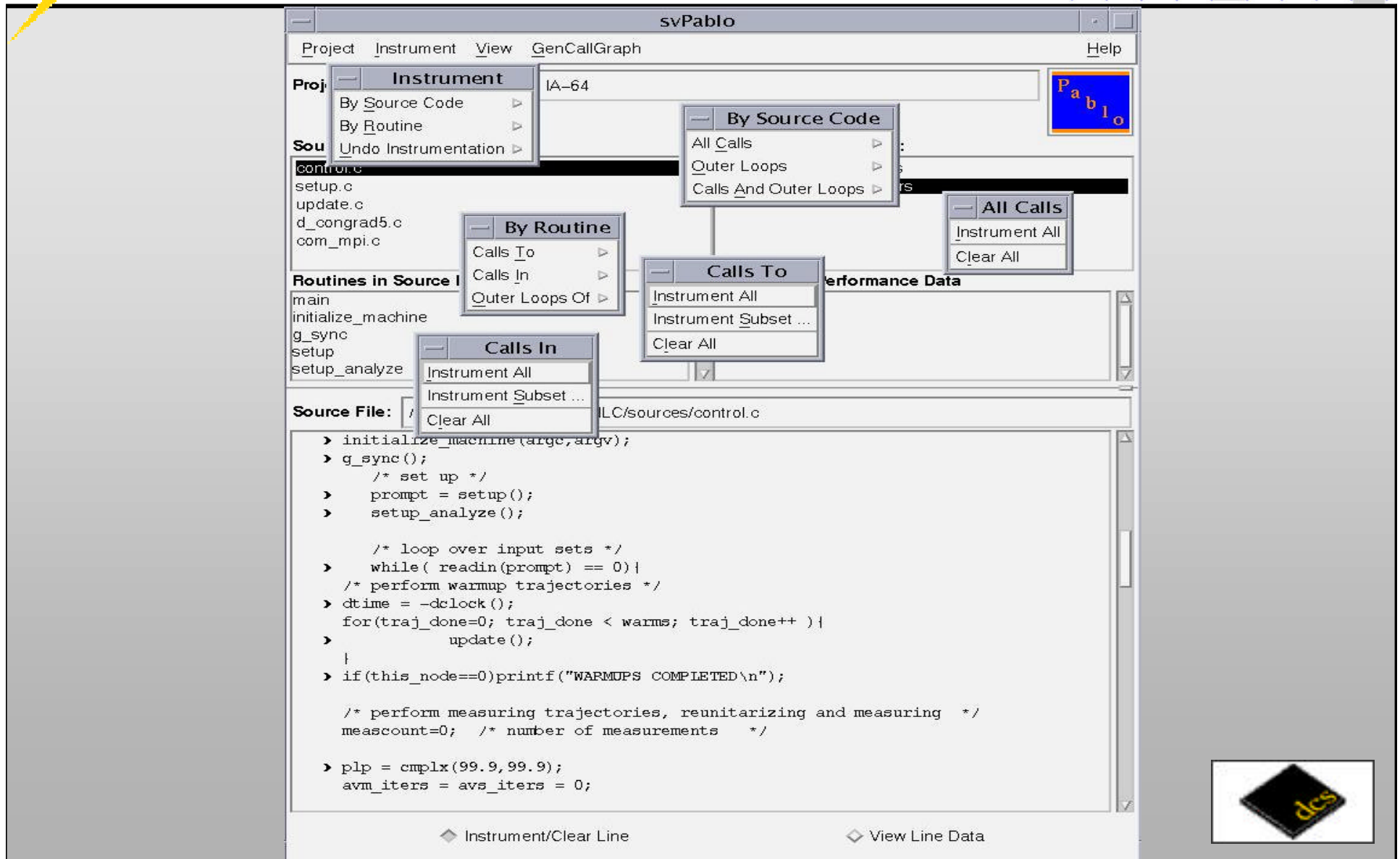

```

/*
 * The following
 * to test some
 */
~> for (i=1; i<=it;
> MPI_Sendrecv(
> MPI_Sendrecv(
    &u[(myend+1)*n], n, MPI_DOUBLE, bottom,
    (myid+1) * blocksize + 1, MPI_COMM_WORLD, &status );
    if (myid == 0) oldNorm = norm;
> MPI_Reduce(&mynorm, &norm, 1, MPI_DOUBLE, MPI_SUM, 0, MPI_COMM_WORLD,
    )
> saveOutput( u, n, i, norm, oldNorm );
MPI_Finalize();
      
```

A dialog box titled "Multiple constructs to instrument on line" is open, showing:

- Currently selected line has these constructs:
 - Outer Loop
 - relax
 - updateOmega
- Type of Instrumentation To Apply:
 - ☒ Instrument
 - ☐ Clear Instrument
- Buttons: Selected, Except, Dismiss, Help

At the bottom of the GUI, there are two buttons: ☒ Instrument/Clear Line and ☒ View Line Data.



The screenshot displays the svPablo software interface, which is used for automatic instrumentation. The interface includes a menu bar with options: Project, Instrument, View, GenCallGraph, and Help. A toolbar contains icons for 'Pablo' and 'dcS'.

The main window is divided into several sections:

- Project:** IA-64
- Instrument:** By Source Code, By Routine, Undo Instrumentation
- Source Files:** control.c, setup.c, update.c, d_congrad5.c, com_mpi.c
- Routines in Source:** main, initialize_machine, g_sync, setup, setup_analyze
- Source File:** /LC/sources/control.c
- Performance Data:** (Empty table)

Four context menus are open, showing options for instrumentation:

- By Source Code:** All Calls, Outer Loops, Calls And Outer Loops
- By Routine:** Calls To, Calls In, Outer Loops Of
- Calls To:** Instrument All, Instrument Subset ..., Clear All
- Calls In:** Instrument All, Instrument Subset ..., Clear All

The source code for `control.c` is displayed in the bottom pane:

```
> initialize_machine(argc,argv);
> g_sync();
  /* set up */
> prompt = setup();
> setup_analyze();

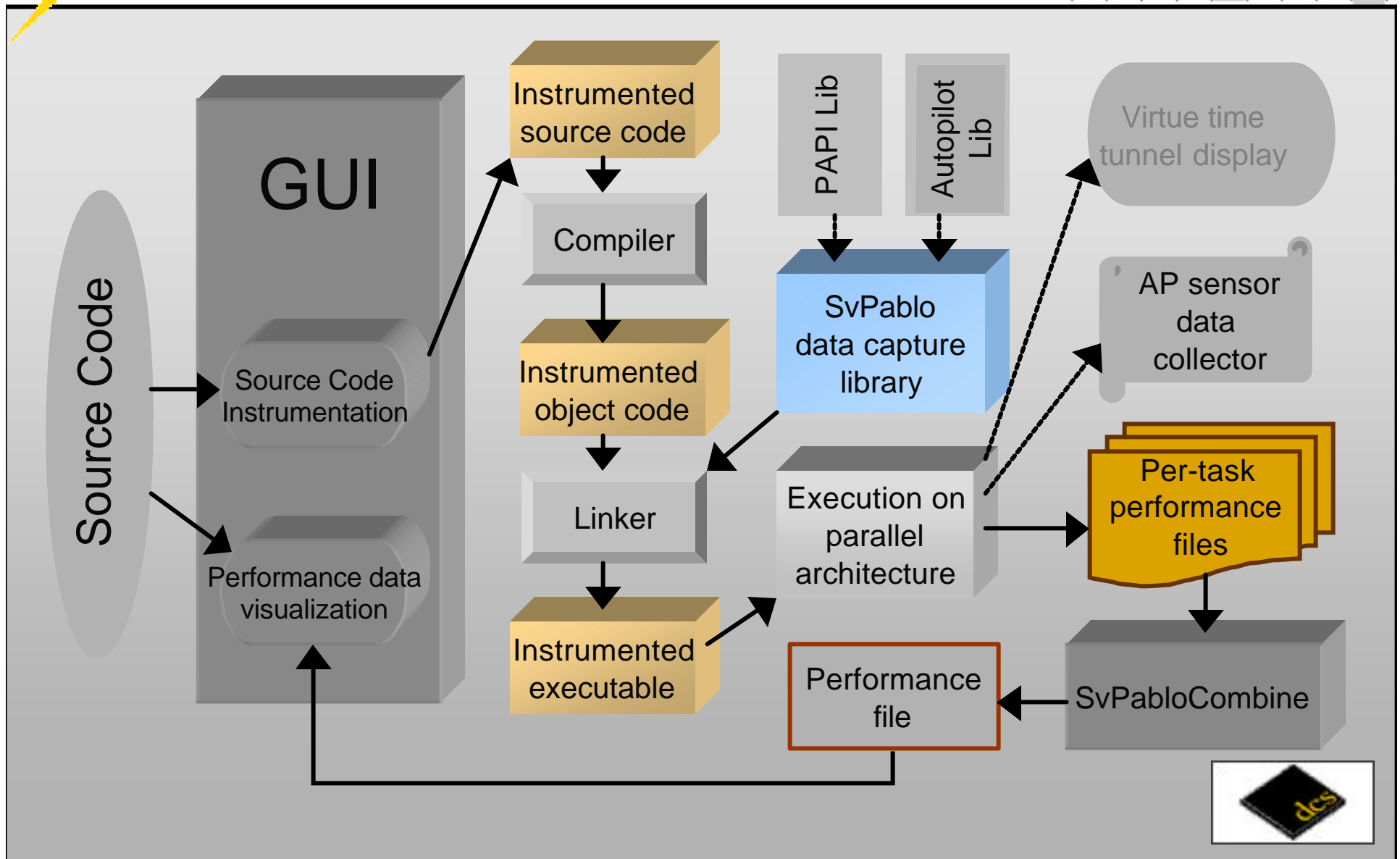
  /* loop over input sets */
> while( readin(prompt) == 0){
  /* perform warmup trajectories */
> dttime = -dclock();
  for(traj_done=0; traj_done < warmups; traj_done++ ){
>     update();
  }
> if(this_node==0)printf("WARMUPS COMPLETED\n");

  /* perform measuring trajectories, reunitarizing and measuring */
  meascount=0; /* number of measurements */

> plp = cmplx(99.9,99.9);
  avm_iters = avs_iters = 0;
```

At the bottom of the interface, there are two buttons: "Instrument/Clear Line" and "View Line Data".

Step 2: Program Compilation and Execution



- Adjust application's Makefile(s)
 - Replace source code filenames
e.g. prog.c ? prog.Context.inst.c
 - Compile *InstrumentationInit.c* and link with it
 - Must always instrument main program
 - Link with `$(SVPABLO)` and `$(PAPI)`

- Execute instrumented executable
- Combine per-task performance files:

```
% SvPabloCombine -o PerfFile c_SDDF*.asc
```



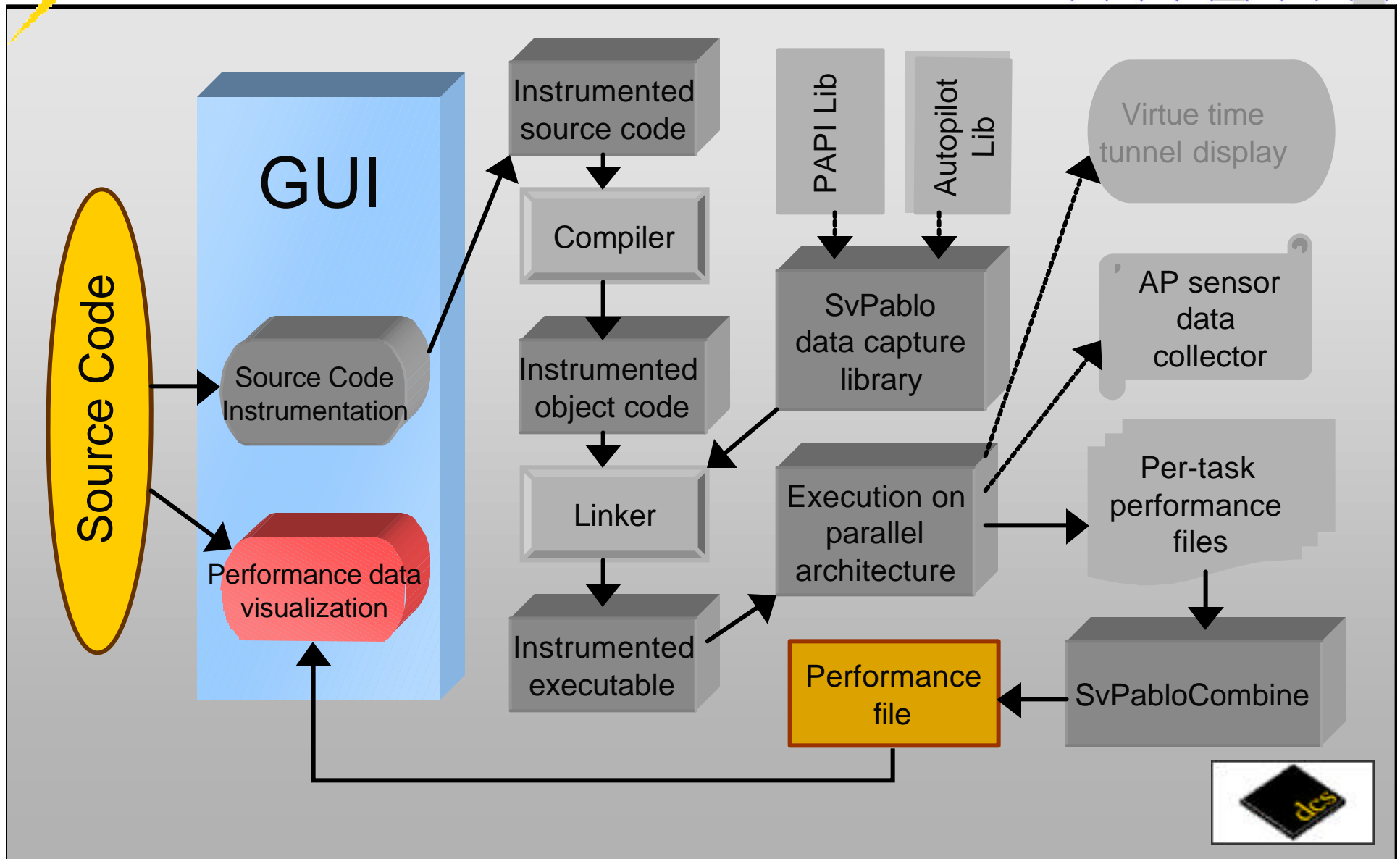
- User-configured file is read at runtime
 - desired PAPI counters are specified in file
 - if file unavailable, default counter set is used
 - SvPabloLibrary versions: with or without PAPI
 - synthesized metrics: e.g. MFLOPS, % branches mispredicted
- 8 hardware counters are available on Seaborg, to see the available HW events, issue the command:

```
% /usr/common/usg/papi/2.1/src/tests/avail
```

See the SvPablo user's guide for details:

<ftp://www-pablo.cs.uiuc.edu/pub/Pablo.Release.5/Documentation/SvPabloGuide.ps.gz>





- Color encoded GUI
 - configurable by the user
 - no change required to display PAPI data
- Aggregate displays
 - mean and standard deviation values across processors
 - maximum value and its processor number
 - minimum value and its processor number
- Detailed displays
 - individual metric values per processor



svPablo

Project Instrument View Help

Project Description: Red Black SOR in C using MPI

Source Files:
prbsor.c
prelax.c
p_io.c

Performance Contexts:
Origin 2000 – 16 R10K processors – 800x800
Power Challenge – 8 R10K Processors – 125x125
NoW – 8 Sun UltraSparcs – 800x800
NoW – 4 Sun UltraSparcs – 125x125
Example: no instrumentation

Routines in Source File

Routines in Performance Data
MPI_Sendrecv
MPI_Send
MPI_Reduce
MPI_Bcast
MPI_Recv

Source File:

Instrument/Clear Line View Line Data

count,
exclusive
duration



Project Instrument View Help

Project Description: Red Black SOR in C using MPI

Source Files:

prbsor.c
prelax.c
p_io.c

Performance Contexts:

Origin 2000 - 16 R10K processors - 800x800
Power Challenge - 8 R10K Processors - 125x125
NoW - 8 Sun UltraSparcs - 800x800
NoW - 4 Sun UltraSparcs - 125x125
Example: no instrumentation

Routines in Source File

relax
fabs
updateOmega
MPI_Sendrecv

Routines in Performance Data

MPI_Bcast
MPI_Recv
relax
updateOmega
malloc

Source File: /home/reed/derose/DemoPablo/Install/SourceFiles/CMPI_RBSOR/prelax.c

```

void relax( n, omega, f, u, norm )
{
    int n;
    double *omega, *norm;
    double *f, *u;

    {
        int i, k, lsw;
        double resid;
        MPI_Status status;

        *norm = 0.0;

        lsw = firstCol;
        for ( i = mystart; i <= myend; i++ )
        {
            for ( k = lsw; k < n-1; k += 2 )
            {
                resid = u[n*(i+1) + k] + u[n*(i-1) + k] +
                    u[n*i + (k+1)] + u[n*i + (k-1)] -
                    4 * u[n*i + k] - f[n*i + k];
                *norm += fabs(resid);
                u[n*i + k] -= *omega * resid / (-4);
            }
            lsw = 3 - lsw;
        }
    }
}

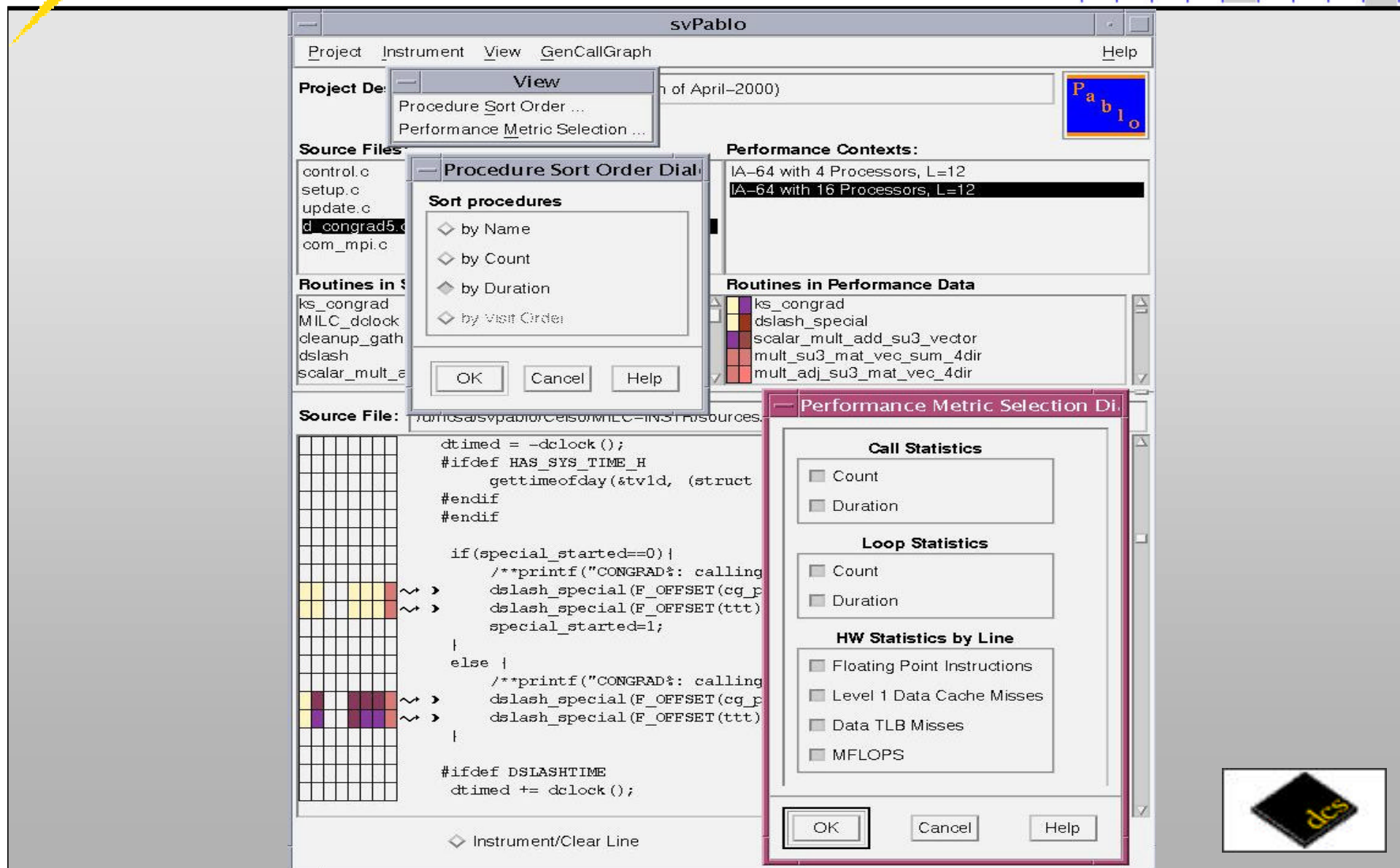
```

Instrument/Clear Line

View Line Data

selected
function



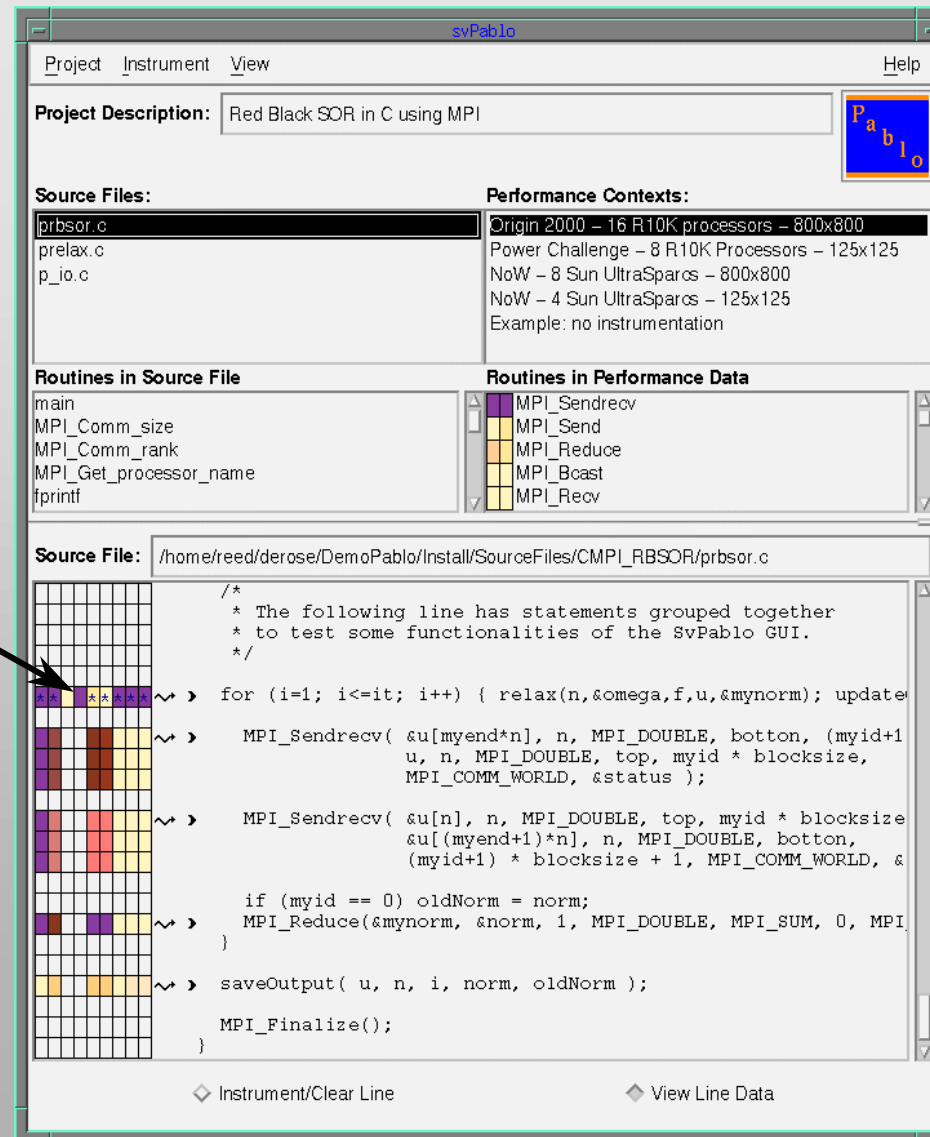


The screenshot shows the **svPablo** application interface. The **View** menu is open, showing options like **Procedure Sort Order...** and **Performance Metric Selection...**. The **Procedure Sort Order Dialog** is also open, showing options to sort procedures by Name, Count, Duration, or Visit Order. The **Performance Metric Selection Dialog** is the primary focus, showing various statistics categories:

- Call Statistics:**
 - ☐ Count
 - ☐ Duration
- Loop Statistics:**
 - ☐ Count
 - ☐ Duration
- HW Statistics by Line:**
 - ☐ Floating Point Instructions
 - ☐ Level 1 Data Cache Misses
 - ☐ Data TLB Misses
 - ☐ MFLOPS

The background shows the **svPablo** main window with a **Project** field, **Source Files** list, **Routines in Performance Data** list, and a **Source File** editor displaying C code. A **Performance Contexts** section shows hardware configurations like "IA-64 with 4 Processors, L=12".

metrics



The screenshot shows the svPablo GUI with the following sections:

- Project Description:** Red Black SOR in C using MPI
- Source Files:** prbsor.c, prelux.c, p_io.c
- Performance Contexts:**
 - Origin 2000 - 16 R10K processors - 800x800
 - Power Challenge - 8 R10K Processors - 125x125
 - NoW - 8 Sun UltraSparcs - 800x800
 - NoW - 4 Sun UltraSparcs - 125x125
 - Example: no instrumentation
- Routines in Source File:** main, MPI_Comm_size, MPI_Comm_rank, MPI_Get_processor_name, fprintf
- Routines in Performance Data:** MPI_Sendrecv, MPI_Send, MPI_Reduce, MPI_Bcast, MPI_Recv
- Source File:** /home/reed/derose/DemoPablo/Install/SourceFiles/CMPI_RBSOR/prbsor.c
- Source Code Visualization:** A grid on the left shows performance metrics for each line of code. The code on the right includes comments and MPI-related functions.

Arrows from the 'metrics' label point to the grid and the code lines.

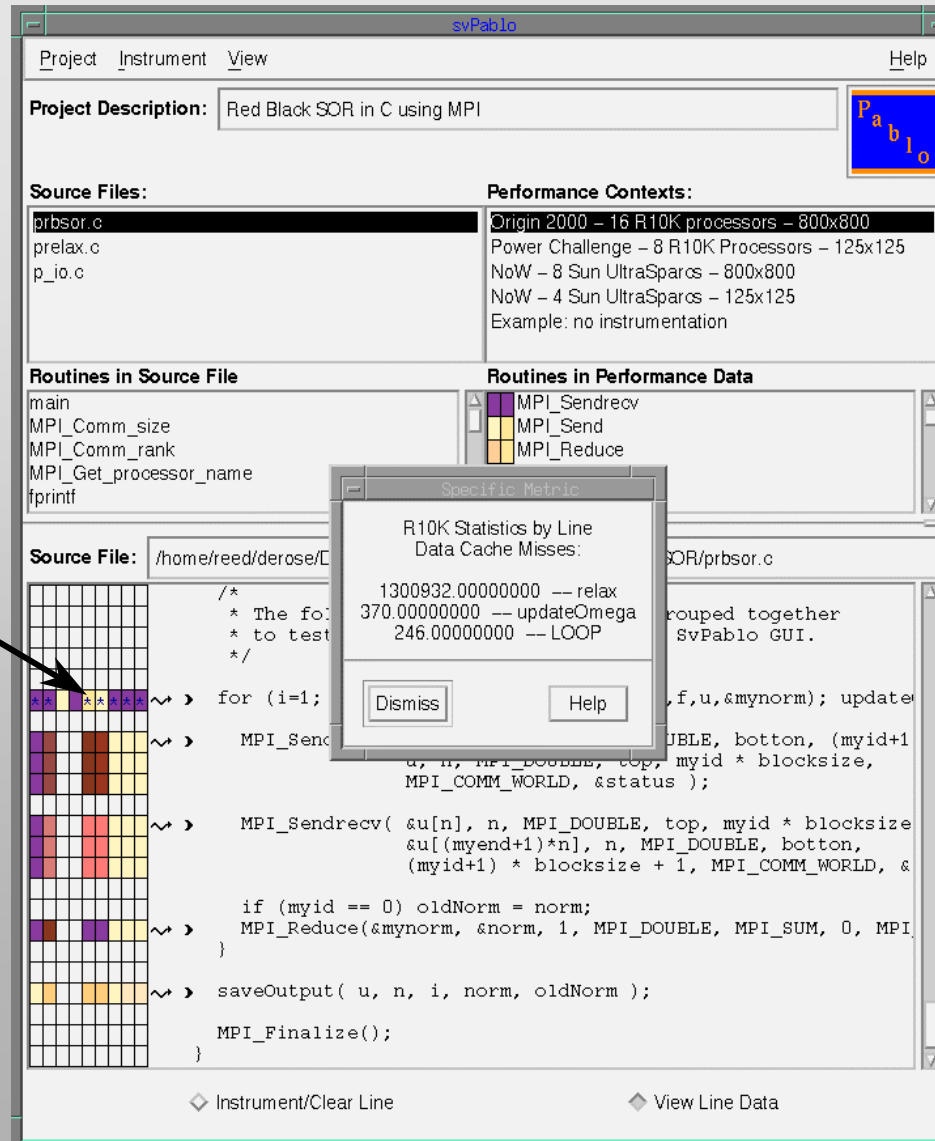
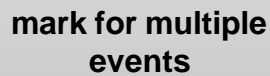


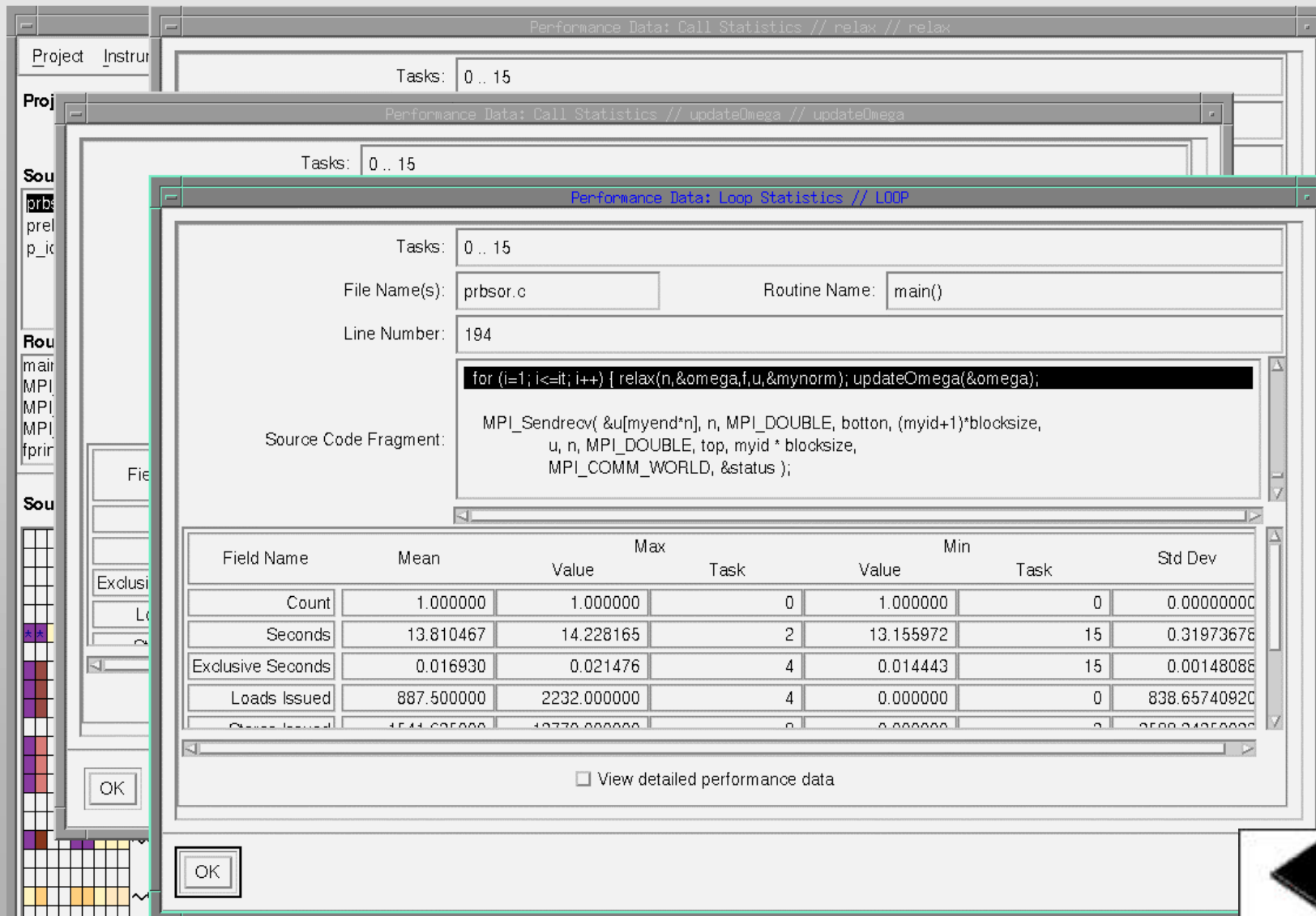


Multiple Events Metric Box



PERC





Currently:

- Tools are being ported to more HPC platforms.
- Additional functionality is being added to the tools.
- Application groups are using the tools and providing feedback.

Future:

- Port tools to all major HPC platforms.
- Improve functionality with the help of feedback from application groups.
- Enhance the functionality of the tools and work on interoperation.
- Integrate performance models with the tools.

✍ The HPC community will be presented with a robust, versatile, and portable suite of performance tools that is suited to a modern HPC environment.

PAPI

- Multi-way multiplexing
- Faster substrates

Dyninst

- Implement full functionality on all platforms
- Build infrastructure for use with parallel applications

Sigma++

- Predict Performance using trace data
- Compile-time instrumentation for collection of data-dependence information

Performance Bounds

- Build on infrastructure for C/C++ and Fortran-77 source code
- Automate the application of performance bounding techniques

SvPablo

- Improve infrastructure for supporting Fortran-90 codes
- Develop an infrastructure for supporting C++ codes
- Add support for OpenMP

Interoperate with the other tools and instrumentation systems:

Dyninst: Use the Dyninst API to allow dynamic instrumentation and analysis at runtime

Performance Bounding: Use to display performance bounds alongside measured performance

Sigma: Enable Sigma instrumentation and visualization of Sigma results